the matrix computation engine includes a lookup table to map first matrix elements having the first size to second matrix elements having the second size, wherein the second matrix elements are provided as the first vector operand.

9. The matrix computation engine as recited in claim 8 further comprising one or more operand memories coupled to the circuit and storing the second vector operand and a third vector operand, wherein the third vector operand has the first matrix elements of the first size, wherein the second vector operand and the third vector operand are read from the one or more operand memories to perform the matrix multiplication operation during use, and wherein the matrix computation engine is configured to map the first matrix elements to generate the first vector operand.

10. The matrix computation engine as recited in claim 9 wherein the lookup table is programmable.

11. The matrix computation engine as recited in claim 10 wherein the lookup table is stored in one of the one or more operand memories.

12. The matrix computation engine as recited in claim 8 wherein the circuit comprises a plurality of multiply-accumulate circuits, wherein the plurality of multiply-accumulate circuits are coupled to receive input matrix elements of the second size and configured to generate result matrix elements of a third size greater than the second size.

13. The matrix computation engine as recited in claim 12 further comprising an output memory, wherein the plurality of multiply-accumulate circuits are coupled to receive matrix elements from the output memory to accumulate with the resulting output vector, wherein the matrix elements from the output memory are of the third size, and wherein the resulting output vector accumulated with the matrix elements from the output memory is written to the output memory.

14. The matrix computation engine as recited in claim 12 wherein a number of the plurality of multiply-accumulate circuits is equal to a number of matrix elements in the output memory.

15. The matrix computation engine as recited in claim 12 wherein a number of the plurality of multiply-accumulate circuits is less than a number of matrix elements in the output vector, and wherein the matrix multiplication operation is completed by iterating the plurality of multiply-accumulate circuits with different portions of the first input operand and the second input operand.

16. A method comprising:
  reading a first vector of first elements from a first operand memory in a computation engine, the first elements having a first size;
  using a lookup table in the computation engine to map each of the elements of the first vector to second elements having a second size greater than the first size, wherein the second elements form a second vector; and
  performing multiply-accumulate operations on the second vector of second elements and a third vector of third elements.

17. The method as recited in claim 16 wherein the first vector, the second vector, and the third vector are vectors of matrices, and wherein the first elements, second elements, and third elements are matrix elements.

18. The method as recited in claim 16 wherein further comprising programming the lookup table from a processor that is coupled to the computation engine.

19. The method as recited in claim 18 further comprising issuing a computation instruction from the processor to the computation engine to cause the multiply-accumulate operations to be performed.

20. The method as recited in claim 18 wherein the lookup table is stored in the first operand memory.

* * * * *